# Demystifying Graph Neural Network Explanations

Anna Himmelhuber, Mitchell Joblin, Martin Ringsquandl, Thomas Runkler

Supported by:

Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag

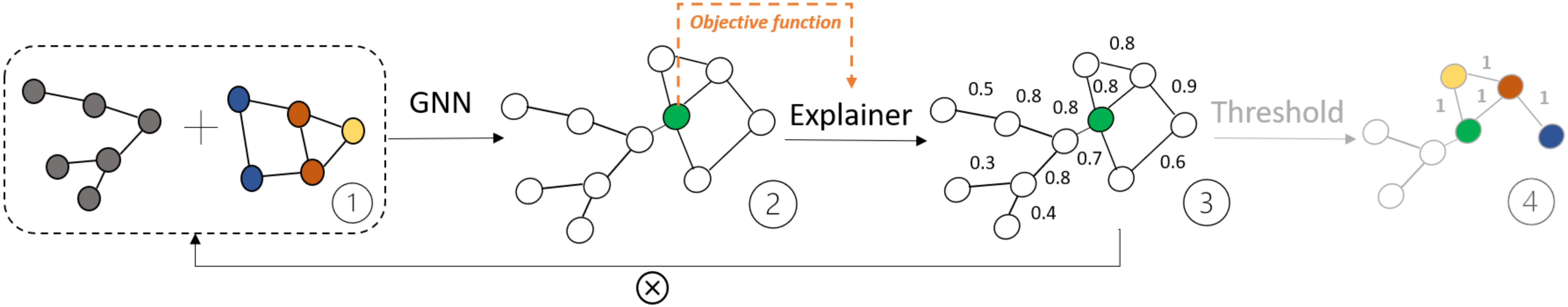
Technische Universität München

SIEMENS

# XAI on Graph Neural Networks

- Many important real-world data sets are graphs or networks
- Graph Neural Networks lack transparency in their decision-making
- Nonetheless, they are a promising candidate for producing reach explanations

# Perturbation-based Explainer Methods




- **GNNExplainer**[1]
- **CF-GNNExplainer**[2]
- **ZORRO**[3]
- **PGExplainer**[4]

[1]Ying, Rex, et al. "Gnnexplainer: Generating explanations for graph neural net-works." Advances in neural information processing systems 32 (2019): 9240
[2]Lucic, Ana, et al. "CF-GNNExplainer: Counterfactual Explanations for GraphNeural Networks."
[3]Funke, Thorben, Megha Khosla, and Avishek Anand. "Hard Masking for Explain-ing Graph Neural Networks." (2020)
[4]. D. Luo, W. Cheng, D. Xu, W. Yu, B. Zong, H. Chen, and X. Zhang, "Parameterizedexplainer for graph neural network," in Advances in neural information processingsystems, 2020.

# Demystifying GNN Explanations

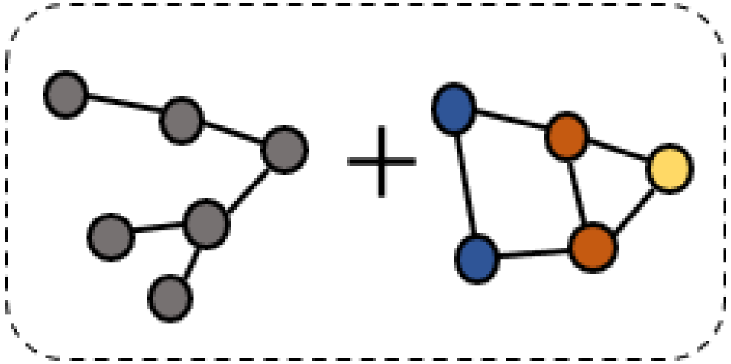
The current methods and data used to evaluate GNN  explanations lack maturity. We explore these existing approaches and identify common pitfalls in three main areas:

- (1) Synthetic data generation process
- (2) Evaluation metrics
- (3) Final presentation of the explanation

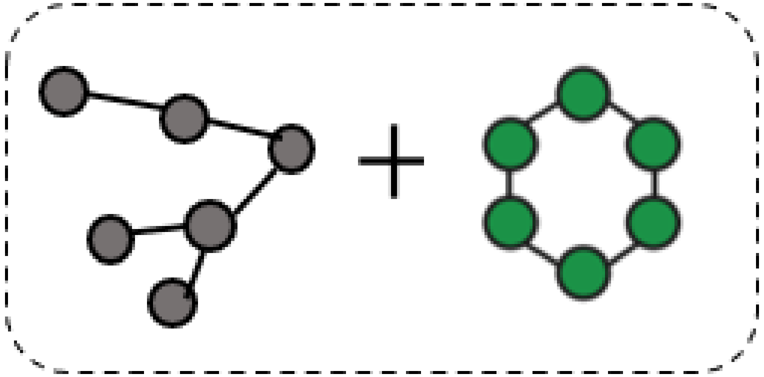
 

# (1) Synthetic data generation process

- Synthetic data sets with intuitive motifs are used for explanation evaluation

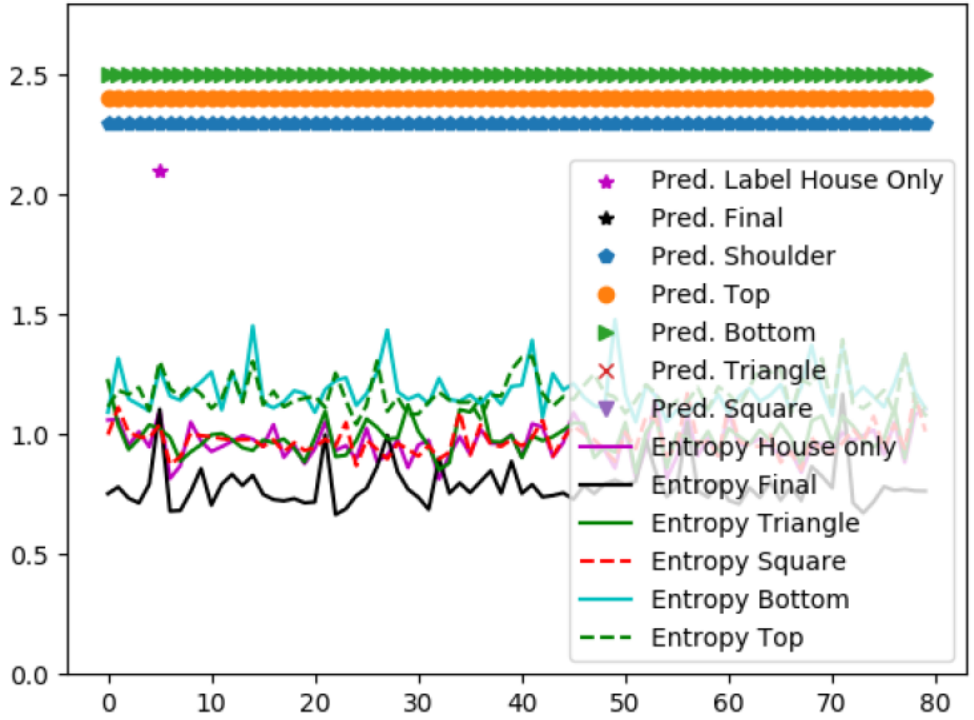


BA Shapes



Tree Cycles

# (1) Synthetic data generation process




Entropy comparison of different motifs for shoulder node

**Remedy: Motif Search**




SIEMENS

# (2) Evaluation metrics

- Differing Accuracy definitions
- ROC AUC overly optimistic for imbalanced data set
- Threshold-dependent metrics should be included

| Class | ROC AUC | SD | PR AUC (proposed) | SD | Recall (proposed) | SD |
|---|---|---|---|---|---|---|
| Top Nodes | 0.98 | 0.07 | 0.69 | 0.19 | 0.65 | 0.18 |
| Shoulder Nodes | **0.98** | 0.91 | **0.51** | 0.19 | 0.51 | 0.13 |
| Bottom Nodes | 0.93 | 0.18 | 0.56 | 0.22 | 0.57 | 0.21 |
| Cycle Nodes | 0.71 | 0.22 | 0.55 | 0.16 | 0.52 | 0.14 |

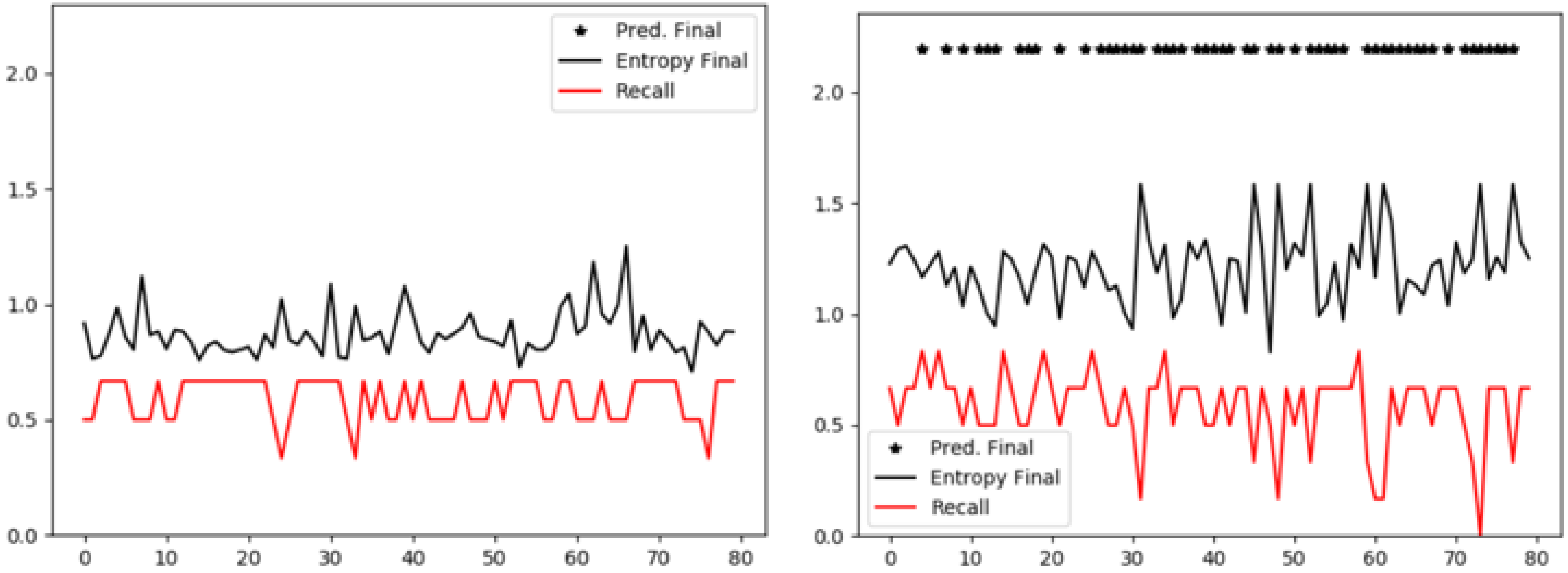
Comparison of ROC AUC, PR AUC and Recall scores

**PR AUC/Recall are more insightful**

# (3) Final presentation of the explanation

- Label flips indicate lacking fidelity of explanations




Label flip for different node classes: Shoulder nodes (left) and top nodes (right)

SIEMENS

# (3) Final presentation of the explanation

- Choice of threshold is non-trivial
- Trade-off between compactness and completeness of explainer subgraph

| Class | Accuracy th = 6 | Accuracy th = 20 |
|---|---|---|
| Top Nodes | 0.65 | 0.98 |
| Shoulder Nodes | 0.51 | 0.82 |
| Bottom Nodes | 0.57 | 0.75 |
| Cycle Nodes | **0.52** | **0.97** |

Recall for different thresholds

**Grid-search to find optimal threshold**




SIEMENS

# References

[1] Arrieta, Alejandro Barredo, et al. ”Explainable Artificial Intelligence (XAI): Con-cepts, taxonomies, opportunities and challenges toward responsible AI.” InformationFusion 58 (2020): 82-115.
[2] Kipf, Thomas N., and Max Welling. ”Semi-supervised classification with graph con-volutional networks.” ICLR (2017).
[3] Lecue, Freddy. ”On the role of knowledge graphs in explainable AI.” Semantic Web11.1 (2020): 41-51.
[4] Lucic, Ana, et al. ”CF-GNNExplainer: Counterfactual Explanations for Graph Neural Networks.” (2021).
[5] Robnik-ˇSikonja, Marko, and Marko Bohanec. ”Perturbation-based explanations ofprediction models.” Human and machine learning. Springer, Cham, 2018. 159-175.
[6] Xu, Keyulu, et al. ”How powerful are graph neural networks?.” ICLR (2018)
[7] Ying, Rex, et al. ”Gnnexplainer: Generating explanations for graph neural net-works.” Advances in neural information processing systems 32 (2019):
[8] Funke, Thorben, Megha Khosla, and Avishek Anand. ”Hard Masking for Explain-ing Graph Neural Networks.” (2020).
[9] Yuan, Hao, et al. ”Explainability in graph neural networks: A taxonomic survey.” (2020).
[10] D. Luo, W. Cheng, D. Xu, W. Yu, B. Zong, H. Chen, and X. Zhang, “Parameterized explainer for graph neural network,” in Advances in neural information processingsystems, 2020.
[11] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji, “On explainability of graph neuralnetworks via subgraph explorations,”

SIEMENS